Comparison of different models for estimating daily yields from a.m./p.m. milkings in Slovenian dairy scheme

J. Jenko, T. Perpar, B. Logar, M. Sadar, B. Ivanovic, J. Jeretina, J. Verbic & P. Podgoršek

Agricultural institute of Slovenia, Hacquetova 17, 1000 Ljubljana, Slovenia

Summary

Evaluation of accuracy and bias for prediction of daily milk yield (DMY), daily fat content (DFC), and daily protein content (DPC) from alternate a.m./p.m. recording scheme was made. The data comprised information on 483 813 test-day records. Five different equations were made from three different models compared on the whole data set or just the upper or the lower quartile of the records. In the first model daily yield or content was included as a dependent variable in regression analysis, whereas in the other two ratio of partial to daily yield or content was dependent variable in regression analysis. The bias on a whole data set was except for one equation, where DPC was underestimated for 0.033%, low. On the upper quartile of data set bias in the underestimation of records was noticed for two equations. In the first equation DMY, DFC, and DPC was on average underestimated for 0.315 kg, 0.196%, and 0.047%, respectively. The third equation on average underestimated DMY, DFC, and DPC for 0.392 kg, 0.211%, and 0.026%, respectively. However the problem with bias at these two equations was also detected in the lower quartile of data set. Here, the records in the first equation for DMY and DFC were on average overestimated for 0.347 kg and 0.170%, respectively and DPC was underestimated for 0.021%. The third equation on average overestimated DMY, DFC, and DPC for 0.329 kg, 0.195%, and 0.023%, respectively. First equation and third equation were both based on the first model. Differences in correlation between estimated and true values were negligible among equations. It was concluded that models where partial to daily yield or content was included as a dependent variable in regression analysis made unbiased estimation, whereas model where daily yield or content was dependent variable in regression analysis made biased estimation with respect to high or low milk yield and fat or protein content.

***Key words**: Alternating a.m./p.m. scheme, Bias, Milk yield, Fat content, Protein content.*

Introduction

Reliable data from milk recording system are important for herd management and genetic improvement in dairy cattle (Liu *et al.*, 2000). Constant pressure of lowering the cost of dairy recording system forces us into implementation of cheaper methods. When we look for a cheaper alternative of A4 recording scheme (monthly records of daily milkings) an implementation of AT recording scheme (alternating morning and evening milking records at subsequent visit) is a logical step. Potential benefits from the adoption of AT recording scheme are the following: more herds can be served by one supervisor, recording costs per cow are lower (Everett and Wadell, 1970a; Everett and Wadell, 1970b; Hargrove and Gilbert, 1984), larger number of young bulls can be involved in progeny testing programs per year, rate of genetic gain is increased (Schaeffer and Rennie, 1976), flexibility in scheduling the work of a supervisor is greater and the method disrupts the milking routine less (Cassandro *et al.*, 1995).

There were several studies conducted to investigate the accuracy of AT recording scheme in comparison with A4 recording scheme (Cassandro *et al.*, 1995; DeLorenzo and Wiggans, 1986; Lee and Wardrop, 1984; Liu *et al.*, 2000; Wiggans, 1981). The model proposed by DeLorenzo and Wiggans (1986) has been widely used in many countries. Here daily yield is estimated from one milking by multiplying the yield from the milking by a factor dependent on the length of interval since the last milking. Factors for estimating daily milk yield (DMY) from partial milk yield (PMY) from morning (a.m.) or evening (p.m.) milking adjusted for milking intervals (MI) can be derived through regression analysis of direct (DMY:PMY) or inverse (PMY:DMY) yield ratios on the length of MI. Cassandro *et al.* (1995) showed that inverse adjustment factors have mathematical features that are more favourable than direct factors. The model proposed by Liu *et al.* (2000) estimates daily yield from single milk testing scheme with a multiple regression method. In contrast with DeLorenzo and Wiggans model (1986) they use daily yield instead of direct or inverse yield rations as the dependent variable in regression analysis. Estimated daily yield is calculated with model where separate regression for every combination of parity, MI, and lactation stage is used. The result is a model with 576 equations for estimating DMY, daily fat yield (DFY), and daily protein yield (DPY).

In March 2004 the AT recording scheme was implemented in Slovenian dairy scheme instead of the ICAR A4 standard reference recording scheme (Sadar *et al.*, 2005). Equations calculated with the model proposed by Klopcic *et al.* (2003) and Klopcic (2004) were used. The main problem observed was systematic underestimation of daily protein content (DPC) (Sadar *et al.*, 2007) and some systematic biases, e.g., underestimation of high daily yields and contents on one hand and overestimation of low daily yields and contents on the other hand (personal communication with breeders). Although Liu *et al.* (2000) warns about some systematic pattern of estimated errors in connection to lactation stage, this problem was also seen in connection to production level. This is quite important for Slovenia where the intensity of milk production differs a lot between farms (Sadar, *et al.*, 2007).

The objectives of this study were to investigate adequacy of official equations proposed for Slovenia and to compare different models for calculation of DMY, daily fat content (DFC), and DPC on the basis of data from single milking.

Material and methods

Data

Milk production data were collected from the database of Slovenian cattle recording scheme (Logar *et al.*, 2005). Combinations of data from regular and supervision dairy recordings carried out from March 2004 through February 2008 were made. With adequate a.m. and p.m. records a calculation of daily yields with regular A4 recording method was made. After screening the records from unreasonable days in milk (DIM) (less than 5 days), MI (less than 540 minutes and more than 870 minutes), more than two times milking per day and problems detected at testing, 483 813 test-day records were used. They were collected from 120 971 lactations of 89 376 cows from 26 046 milk tests in 5 051 herds with all traits recorded in a.m./p.m. or p.m./a.m. milkings. Variables included in the study were: DMY, DPC, DFC, PMY from a.m. or p.m. milking, partial fat content (PFC) from a.m. or p.m. milking, partial protein content (PPC) from a.m. or p.m. milking, MI and DIM.

During the night MI was on average 26 minutes longer compared to daily interval (Table1). This inequality is the main cause of mean difference between a.m. milk yield and p.m. milk yield + 0.40 kg. Slightly higher values for DFC and DPC were observed in p.m. milking. On the other hand DFY and DPY are higher in a.m. milking. The reason is in higher a.m. milk yield in comparison with p.m. milk yield.

Table 1. Descriptive statistics of milk yield and milk composition.

Trait	Number of records	Daily		p.m.		a.m.	
		Mean	Std	Mean	Std	Mean	Std
DMY (kg)[1]	483 813	18.06	7.12	8.83	3.56	9.23	3.76
p.m. : DMY[2]	241 914			0.49	0.05		
a.m. : DMY[3]	241 899					0.51	0.05
DFY (kg)[4]	483 813	0.74	0.29	0.37	0.15	0.38	0.16
DFC (%)[5]	483 813	4.18	0.71	4.22	0.82	4.15	0.82
p.m. : DFY[6]	241 914			0.49	0.07		
a.m. : DFY[7]	241 899					0.51	0.07
DPY (kg)[8]	483 813	0.61	0.22	0.30	0.11	0.31	0.12
DPC (%)[9]	483 813	3.41	0.38	3.43	0.39	3.40	0.39
p.m. : DPY[10]	241 914			0.49	0.05		
a.m. : DPY[11]	241 899					0.51	0.05
MI (min)[12]	483 813	1436.91	20.57	705.33	44.39	731.55	46.69

[1]Daily milk yield.
[2]Ratio of evening milk yield to daily milk yield
[3]Ratio of morning milk yield to daily milk yield
[4]Daily fat yield.
[5]Daily fat content.
[6]Ratio of evening fat yield to daily fat yield
[7]Ratio of morning fat yield to daily fat yield
[8]Daily protein yield.
[9]Daily protein content.
[10]Ratio of evening protein yield to daily protein yield
[11]Ratio of morning protein yield to daily protein yield
[12]Milking interval.

Models

In the study, five equations were analysed. They were made on three different models. Model 1 was proposed by Klopcic *et al.* (2003) and Klopcic (2004), Model 2 by DeLorenzo and Wiggans (1986) and an alternative Model 3 which is a combination of Model 1 and Model 2 mentioned previously.

Model 1: The following formula was used for estimating DMY, DFC, and DPC with model proposed by Klopcic *et al.* (2003) and Klopcic (2004):

$$y_i = \mu + b_1 * m_i + b_2 * t_i + e_i \quad (1)$$

where y is an estimated DMY, DFC or DPC for i measurement, μ is the estimated intercept for a.m. or p.m. milking, b_1 and b_2 are the estimated regression coefficient for a.m. or p.m. milking, m_i is the PMY, PFC or PPC, t is the MI in minutes, and e_i is the residual

Model 2: For the model developed by DeLorenzo and Wiggans (1986) the factors proposed by ICAR (2007) were used. Here DMY and DFC estimates are based on measured yield or content and milking frequency. An adjustment factor accounts for differences in the average MI between the preceding milking and the time of measured milking. An additional adjustment is applied to milk yield for the interaction between MI and stage of lactation with mid lactation 158 DIM set to zero. DPY is calculated from the measured percentage and the adjusted milk yield. The MI is divided into 15-minutes classes and formulas for prediction sample day yields and percentages in herds with two milkings are:

$$DMY = FM * PMY + c * (DIM - 158) \quad (2)$$

$$DFC = FF * PFC \text{ ; } DFY = DMY * DFC \text{ ; } DPY = DMY * DPC \quad (3)$$

where *FM* is the factor for milk yield, c is the covariate, and *FF* is the factor for fat content.

Model 3: A new model which is a combination of DeLorenzo and Wiggans model (1986) and model by Klopcic *et al.* (2003) and Klopcic (2004) was proposed. In this model multiple regression method that calculates factors for estimating DMY, DFY, and DPY is used. Here an inverse ratio of partial to daily yield from a single milking is included as a dependent variable in regression analysis. The formula is applied separately to partial daily yields from a.m. or p.m. milking. For a given trait one formula is to be estimated for calculating daily yield based on partial yield from either a.m. or p.m. milking. There were two proposals of this model tested:

$$y_i = \mu + b_1 * t_i + e_i \quad (4)$$

and

$$y_i = \mu + b_1 * t_i + b_2 * t_i^2 + e_i \quad (5)$$

where y is an estimated factor (inverse ratio of partial to daily yield) for calculation of DMY, DFY, and DPY for i measurement, μ is the estimated intercept for a.m. or p.m. milking, b_1 and b_2 are the estimated regression coefficient for a.m. or p.m. milking, t is the MI in minutes, and e is the residual.

Estimated DMY, DFY, and DPY are calculated with the formulas:

$$DMY = PMY / FM \, ; \quad DFY = PFY / FF \, ; \quad DPY = PPY / FP \qquad (6)$$

where FM is the estimated factor for milk yield, FF is the estimated factor for fat yield and FP is the estimated factor for protein yield.

Component percentages of fat and protein on a daily basis are then calculated by dividing estimated DFY or DPY by estimated DMY:

$$DFC = (DFY / DMY) * 100 \, ; \quad DPC = (DPY / DMY) * 100 \qquad (7)$$

In the analysis we compare five equations, where the first one (E1) uses the coefficients proposed by Klopcic *et al.* (2003) and Klopcic (2004), and the second one (E2) factors proposed by ICAR (2007) for the model by DeLorenzo and Wiggans (1986). The third (E3), fourth (E4) and fifth (E5) equations have factors calculated on our data set, whereas E3 is based on Model 1 and E4 and E5 on Model 3. Equation E4 is based on MI as linear and E5 as quadratic effect. For the calculation of adjustment factors R statistics package (R Development Core Team, 2008) was used. LM procedure was used for the estimation of coefficients for DMY, DFC, and DPC in Model 1 and factors for DMY, DFY, and DPY in Model 3.

Equations were evaluated on the basis of correlation between the estimated and measured DMY, DFC, and DPC and on the basis of the differences between predicted and true DMY, DFC, and DPC. To study the problem of underestimation of high DMY, DFC, and DPC and overestimation of low DMY, DFC, and DPC upper and lower quartile of records per each trait were analysed.

Results

The accuracy of DMY, DFC, and DPC predicted from analysed equations is shown in Table 2. There is actually no difference in correlation coefficient between the selected equations. Except in the equation E1 for DPC the bias is quite low. In the equation E1 is on average DPC underestimated for 0.033%. This means that equations officially proposed for Slovenia made a systematic bias which on average underestimates the true DPC. From the results shown in Table 2 it cannot be concluded which model estimates the daily milk records better.

The accuracy to predict DMY, DFC, and DPC on the basis of the upper quartile data sets is shown in Table 3. The boundary is set to > 22.4 kg for DMY, > 4.62% for DFC, and > 3.66% for DPC. The correlations for this part of data set are lower compared with the whole data set (Table 2). There are almost no differences in correlation coefficients between equations. However, the differences in bias can be clearly seen. Equations E1 and E3 systematically underestimate DMY, DFC, and DPC. Equation E1 underestimates DMY, DFC, and DPC for 0.315 kg, 0.196%, and 0.047%, respectively. Equation E3 underestimates DMY, DFC, and DPC for 0.392 kg, 0.211%, and 0.026%, respectively. On the other hand, equations E2, E4, and E5 are quite unbiased and make a good estimation of daily yield and composition. The problem of underestimation of high daily yield and contents is solved with Model 2 and Model 3 on which equations E2, E4, and E5 are based.

Table 2. Accuracy and bias of DMY, DFC, and DPC on the whole data set.

Trait	Equation	r[4]	Bias[5]	Std[6]	Min[7]	Max[8]
DMY (kg)[1]	E1	0.979	0.044	1.444	-18.31	18.11
	E2	0.980	-0.048	1.445	-19.18	19.05
	E3	0.980	-0.002	1.428	-18.40	18.66
	E4	0.980	-0.002	1.431	-19.14	19.51
	E5	0.980	-0.001	1.431	-19.14	19.50
DFC (%)[2]	E1	0.878	-0.006	0.342	-3.43	2.73
	E2	0.879	0.001	0.387	-3.86	3.53
	E3	0.879	0.000	0.340	-3.35	2.57
	E4	0.879	-0.007	0.385	-3.88	3.64
	E5	0.879	-0.003	0.385	-3.88	3.66
DPC (%)[3]	E1	0.975	-0.033	0.087	-2.38	1.73
	E2	0.974	0.001	0.089	-2.32	1.83
	E3	0.975	0.000	0.086	-2.32	1.71
	E4	0.975	0.000	0.088	-2.34	1.81
	E5	0.975	0.000	0.088	-2.34	1.81

[1]Daily milk yield.
[2]Daily fat content.
[3]Daily protein content.
[4]Correlation coefficient between the estimated and measured DMY, DFC, and DPC.
[5]Difference between an estimator's expected value and the true value of the parameter being estimated.
[6]Standard deviation.
[7]Minimum difference between an estimator's expected value and the true value of the parameter being estimated.
[8]Maximum difference between an estimator's expected value and the true value of the parameter being estimated.

Table 3. Accuracy and bias of DMY, DFC, and DPC, on upper 25% records from sample data.

Trait	Equation	r[4]	Bias[5]	Std[6]	Min[7]	Max[8]
DMY (kg)[1]	E1	0.923	-0.315	1.867	-18.16	17.63
	E2	0.927	0.080	1.865	-19.18	17.75
	E3	0.925	-0.392	1.836	-18.40	16.91
	E4	0.928	-0.014	1.859	-19.14	17.72
	E5	0.928	-0.011	1.858	-19.14	17.74
DFC (%)[2]	E1	0.709	-0.196	0.365	-3.43	2.73
	E2	0.711	-0.002	0.439	-3.86	3.53
	E3	0.710	-0.211	0.356	-3.35	2.57
	E4	0.709	-0.012	0.441	-3.88	3.64
	E5	0.709	-0.007	0.441	-3.88	3.66
DPC (%)[3]	E1	0.920	-0.047	0.098	-2.38	1.50
	E2	0.920	0.001	0.101	-2.32	1.60
	E3	0.921	-0.026	0.096	-2.32	1.48
	E4	0.921	0.000	0.100	-2.34	1.58
	E5	0.921	0.000	0.100	-2.34	1.58

[1, 2, 3, 4, 5, 6, 7, 8]For description see Table 2.

The accuracy of DMY, DFC, and DPC prediction on the basis of the lower quartile data set is shown in Table 4. The boundary is set to < 12.9 kg for DMY, < 3.71% for DFC, and < 3.14% for DPC. Correlation coefficients are lower than in case of the upper quartile (Table 3) indicating that in cows given small amount of milk the accuracy of prediction is lower than in cows with high DMY. Differences in correlation coefficients between equations were small (Table 4), but on the other hand there were considerable differences in bias. Equation E1 overestimates DMY and DFC for 0.347 kg and 0.170%, respectively and underestimates DPC for 0.021%. Equation E3 overestimates DMY, DFC, and DPC for 0.329 kg, 0.195%, and 0.023%, respectively. The bias is much lower in equations E2, E4, and E5 and they seem to be more reliable for prediction of DMY, DFC, and DPC on the lower quartile of data set. As in the upper quartile, here too may be observed that the problem of overestimation of low daily yield and contents is solved with Model 2 and Model 3, on whose equations E2, E4, and E5 are based.

Table 4. Accuracy and bias of DMY, DFC, and DPC on lower 25% records from sample data.

Trait	Equation	r[4]	Bias[5]	Std[6]	Min[7]	Max[8]
DMY (kg)[1]	E1	0.892	0.347	1.076	-9.22	9.61
	E2	0.902	-0.024	1.051	-8.90	10.15
	E3	0.897	0.329	1.043	-8.80	9.71
	E4	0.908	0.004	1.017	-8.88	10.16
	E5	0.908	0.003	1.016	-8.88	10.16
DFC (%)[2]	E1	0.654	0.170	0.310	-1.72	2.52
	E2	0.653	0.002	0.381	-2.11	3.07
	E3	0.656	0.195	0.302	-1.61	2.52
	E4	0.657	-0.004	0.376	-2.12	3.05
	E5	0.657	-0.000	0.377	-2.11	3.06
DPC (%)[3]	E1	0.901	-0.021	0.070	-1.09	1.01
	E2	0.900	0.001	0.072	-1.12	1.09
	E3	0.902	0.023	0.068	-1.03	1.05
	E4	0.902	0.001	0.071	-1.11	1.07
	E5	0.902	0.001	0.071	-1.11	1.07

[1, 2, 3, 4, 5, 6, 7, 8] For description see Table 2.

Conclusion

This study showed that bias from DMY, DFC, and DPC which is related to high or low milk yield and fat or protein content can be removed from the model for estimating dairy recordings from a.m./p.m. milkings. The problem can be solved with the Model 2 and Model 3. The problem in the equation E1 made from Model 1 is a systematic underestimation of DPC, which is not seen in any other model. Correlation between estimated and predicted value differs with respect to change in DMY, DFC, and DPC. They are lower for the upper and the lower quartile of records; this means that the accuracy of estimation is not so high for this part of data.

Alternative model which was proposed in this study become an official model for estimating daily yields from a.m./p.m. milkings in Slovenian dairy scheme after the 1st of May 2008. Here, an inverse ratio of partial to daily yield from a single milking is included as a dependent variable in regression analysis with respect to time between milkings.

List of references

Cassandro, M., P. Carnier, L. Gallo, R. Mantovani, B. Contiero, G. Bittante and G.B. Jansen. 1995. Bias and accuracy of single milking testing schemes to estimate daily and lactation milk yield. J. Dairy Sci. 78: 2884-2893.

DeLorenzo, M.A. and G.R. Wiggans. 1986. Factors for estimating daily yield of milk, fat, and protein from a single milking for herds milked twice a day. J. Dairy Sci. 69: 2386-2394.

Everett, R.W. and L.H. Wadell. 1970a. Sources of variation affecting the difference between morning and evening daily milk production. J. Dairy Sci. 53: 1424-1429.

Everett, R.W. and L.H. Wadell. 1970b. Sources of variation affecting ratio factors for estimating total daily milk yield from individual milkings. J. Dairy Sci. 53: 1430-1435.

Hargrove, G.L. and G.R. Gilbert, 1984. Differences in morning and evening sample milkings and adjustment to daily weights and percents. J. Dairy Sci. 67: 194-200.

Klopcic , M., Š. Malovrh, G. Gorjanc, M. Kovac and J. Osterc. 2003. Prediction of daily milk fat and protein content using alternating (AT) recording scheme. Czech J. Anim. Sci. 48: 449-458.

Klopcic , M. 2004. Optimizacija vrednotenja proizvodnosti krav v mleèni usmeritvi [Optimization of milk recording practices in dairy cows]. PhD thesis. University of Ljubljana, Biotechnical faculty, Zootechnical Department, Ljubljana, 171 pp.

ICAR. 2007. ICAR Guidelines approved by the general assembly held in Kuopio, Finland on 9 June 2006. Rome, pp. 475.

Lee, A. J. and J. Wardrop. 1984. Predicting daily milk yield, fat percent, and protein percent from morning or afternoon tests. J. Dairy Sci. 67: 351-360.

Liu, Z., R. Reents, F. Reinhardt and K. Kuwan. 2000. Approaches to estimating daily yield from single milk testing schemes and use of a.m.-p.m. records in test-day model genetic evaluation in dairy cattle. J. Dairy Sci. 83: 2672-2682.

Logar, B., P. Podgoršek, J. Jeretina, B. Ivanovic and T. Perpar. 2005. On-line available milk-recording data for efficient support of farm management. In: Kuipers, A., M. Klopcic, and T. Cled (Editors), Knowledge transfer in cattle husbandry: new management practices, attitudes and adaptation. Wageningen Academic Publishers, Wageningen, the Netherlands, EAAP Series 117: 227-230.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <*www.R-project.org*>.

Sadar, M., P. Podgoršek, T. Perpar, B. Logar, B. Ivanovic, J. Jeretina, M. Prevolnik. 2005. Rezultati kontrole prireje mleka in mesa Slovenija 2004 [Results of animal recording, Slovenia 2004]. Agricultural institute of Slovenia, Ljubljana, pp. 50.

Sadar, M., A. Biderman, J. Jeretina, B. Logar, T. Perpar, P. Podgoršek and B. Ivanovic. 2007. Rezultati kontrole prireje mleka in mesa Slovenija 2006 [Results of animal recording, Slovenia 2006]. Agricultural institute of Slovenia, Ljubljana, pp. 56.

Schaeffer, L. R. and J. C. Rennie. 1976. AM-PM testing for estimating lactation yields. Can. J. Anim. Sci. 56: 9-15.

Wiggans, G.R. 1981. Methods to estimate milk and fat yields from a.m./p.m. plans. J. Dairy Sci. 64: 1621-1624.